Assimilating near-real-time mass balance stake readings into a model ensemble using a particle filter

Johannes Marian Landmann¹,², Hans Rudolf Künsch³, Matthias Huss¹,²,⁴, Christophe Ogier¹,², Markus Kalisch³, and Daniel Farinotti¹,²

¹Laboratory of Hydraulics, Hydrology and Glaciology (V AW), ETH Zurich, Zurich, Switzerland
²Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland
³Seminar for Statistics, ETH Zurich, Zurich, Switzerland
⁴Department of Geosciences, University of Fribourg, Fribourg, Switzerland

Correspondence: Johannes Marian Landmann (landmann@vaw.baug.ethz.ch)

Received: 25 September 2020 – Discussion started: 4 November 2020
Revised: 15 August 2021 – Accepted: 5 October 2021 – Published: 1 November 2021

Abstract. Short-term glacier variations can be important for water supplies or hydropower production, and glaciers are important indicators of climate change. This is why the interest in near-real-time mass balance nowcasting is considerable. Here, we address this interest and provide an evaluation of continuous observations of point mass balance based on online cameras transmitting images every 20 min. The cameras were installed on three Swiss glaciers during summer 2019, provided 352 near-real-time point mass balances in total, and revealed melt rates of up to 0.12 m water equivalent per day (m w.e. d⁻¹) and of more than 5 m w.e. in 81 d. By means of a particle filter, these observations are assimilated into an ensemble of three TI (temperature index) and one simplified energy-balance mass balance models. State augmentation with model parameters is used to assign temporally varying weights to individual models. We analyze model performance over the observation period and find that the probability for a given model to be preferred by our procedure is 39 % for an enhanced TI model, 24 % for a simple TI model, 23 %, for a simplified energy balance model, and 14 % for a model employing both air temperature and potential solar irradiation. When compared to reference forecasts produced with both mean model parameters and parameters tuned on single mass balance observations, the particle filter performs about equally well on the daily scale but outperforms predictions of cumulative mass balance by 95 %–96 %. A leave-one-out cross-validation on the individual glaciers shows that the particle filter is also able to reproduce point observations at locations not used for model calibration. Indeed, the predicted mass balances is always within 9 % of the observations. A comparison with glacier-wide annual mass balances involving additional measurements distributed over the entire glacier mostly shows very good agreement, with deviations of 0.02, 0.07, and 0.24 m w.e.

1 Introduction

Glaciers around the world are shrinking. For example, Switzerland has lost already more than a third of its glacier volume since the 1970s (Fischer et al., 2015), glaciers are currently melting at about −0.6 m w.e. a⁻¹ (water equivalent per year) on average (Sommer et al., 2020), and it is expected that glaciers will continue to lose mass (Jouvet et al., 2011; Salzmann et al., 2012; Beniston et al., 2018; Zekollari et al., 2019). Since glaciers are important for the supply of drinking water or for irrigation and electricity production, there is high interest in near-real-time glacier mass balance information. Such information has also become important in the context of public outreach, e.g., for demonstrating the consequences of climate change (e.g., Euronews, 2019; Science Magazine, 2019).

A glacier mass balance nowcasting framework assimilating relevant observations could deliver these near-real-time mass balances whenever required. While nowcasting frameworks exist, for example, for the mass balance of the Greenland Ice Sheet (Fettweis et al., 2013; NSIDC, 2020a), for snow (NSIDC, 2020b; SLF, 2020) or for hydrological
In many cases, mass balance analyses are available twice a year and are based on seasonal in situ observations (Cogley et al., 2011). This relatively low frequency is related to the fact that in situ observations are expensive in terms of both time and manpower. Only recently have low-cost and high-frequency monitoring approaches emerged (Hulth, 2010; Fausto et al., 2012; Keeler and Brugger, 2012; Biron and Rabatel, 2019; Carturan et al., 2019; Gugerli et al., 2019; Netto and Arigony-Neto, 2019). However, even with these observations, it is not straightforward to provide analyses at higher frequencies. This is because near-real-time estimates are often based on ensemble modeling in order to enable a correct quantification of uncertainties. Ensemble modeling is used in glaciology in the context of model intercomparison projects (Hock et al., 2019), future projections for ice sheets and mountain glaciers (e.g., Ritz et al., 2015; Shannon et al., 2019; Golledge, 2020; Marzeion et al., 2020; Seroussi et al., 2020), and also determining the initial conditions for modeling (Eis et al., 2019). However, ensembles are currently not prominent in the calculation of seasonal or daily glacier mass balances. Another reason why calculating higher-frequency glacier mass balance analyses is not straightforward is the lack of knowledge about the short-term variability in the parameters of the necessary models. Temperature index (TI) models, for example, are parametrizations of the full energy balance equation, and they offset some of the changes occurring in the driving processes through parameter fluctuations (Ohmura, 2001; Lang and Braun, 1990; Hock, 2003; Hock et al., 2005). In a comparison of four TI models and a full energy balance model, Gabbi et al. (2014) showed that all models perform very similarly on a multi-year scale.

In this study, we address the issue of low-frequency observations, ensemble modeling, and the lack of knowledge about short-term parameter variability as part of the project CRAMPON (Cryospheric Monitoring and Prediction Online). The latter aims at delivering near-real-time glacier mass balance estimates for mountain glaciers using data assimilation. To obtain high-frequency data at a relatively low cost, we equipped three Swiss glaciers – Glacier de la Plaine Morte, Findelgletscher, and Rhonegletscher – with seven cameras in total. The cameras were operated in summer 2019 and took images of a mass balance stake marked every 2 cm at 20 min intervals, thus providing estimates of surface point mass balance aggregated to the daily scale. By using a particle filter (e.g., Arulampalam et al., 2002; Beven, 2009; Magnusson et al., 2017), we assimilate these observations into an ensemble of three TI models and one simplified energy balance model.

Ensemble stability and suitability for operational use is ensured by designing the particle filter such that, at any instance, each model has a minimum contribution to the mass balance model ensemble. In particular, models with temporarily bad performance are not excluded from the ensemble and can thus regain weight later. To address parameter uncertainty, we drive the mass balance model ensemble with both Monte Carlo samples of uncertain meteorological input and prior parameter distributions obtained from calibration to past, longer-term seasonal mass balance series. By using an augmented state formulation of the particle filter, we constrain model parameters as well (e.g., Ruiz et al., 2013). We are not aware of glacier mass balance studies that have applied a multi-model ensemble based on a particle filter with the resampling methods we propose, although multi-model particle filters have been used for other applications (e.g., Kreucher et al., 2004; Ristic et al., 2004; Saucan et al., 2013; Wang et al., 2016).

As a result, we demonstrate (1) how a workflow including daily melt observations, ensemble modeling, and data assimilation works in practice, (2) to what extent the assimilated mass balances are able to reproduce cumulative observations, and (3) how the ensemble performs with respect to both reference forecasts and seasonal analyses from in situ measurements.

2 Study sites, data, and field instrumentation

We use Glacier de la Plaine Morte, Rhonegletscher, and Findelgletscher in summer 2019 as test sites (Fig. 1). The basic morphological characteristics and instrumentations of these glaciers are given in Table 1.

2.1 Continuous in situ mass balance observations

2.1.1 Technical camera station setup

For acquiring daily point mass balances in the field, we use off-the-shelf cameras and logger boxes from the company Holfly Ltd. We mount the cameras to a construction of aluminum stakes that we designed for glacier applications. Figure 2 provides an overview of the camera installation.

The camera observes an ablation stake, which is marked with colored tape at 2 cm intervals. When the surface melts, the aluminum construction holding the camera slides along the mass balance stake. Pictures are taken every 20 min and are sent in real-time to our servers via the Swiss mobile phone network. Differences between subsequent pictures are used to infer daily glacier surface height changes relative to the stake top, which are the basis for ablation measurements (Cogley et al., 2011). All pieces of the construction
Figure 1. (a) Locations of the glaciers equipped with cameras within Switzerland and (b–d) detailed topographic maps of the glaciers with dots for cameras (red) and reference mass balance stakes (blue). Coordinates are given as Swiss coordinates (EPSG:21781). The blue glacier outlines stem from GLAMOS (Glacier Monitoring Switzerland), and background web mapping service tiles are provided by © swisstopo and © Google Maps.

Table 1. Main characteristics and installed cameras for the investigated glaciers. Glacier area and elevation range refer to the year 2019 (GLAMOS, 2020), and slope and aspect have been calculated using a recent DEM (digital elevation model) (swisstopo, 2020).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Glacier de la Plaine Morte</th>
<th>Findelgletscher</th>
<th>Rhonegletscher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (km²)</td>
<td>7.1</td>
<td>12.7</td>
<td>15.3</td>
</tr>
<tr>
<td>Elevation range (m a.s.l.)</td>
<td>2470–2828</td>
<td>2561–3937</td>
<td>2223–3596</td>
</tr>
<tr>
<td>Average slope (°)</td>
<td>6</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Average aspect (°)</td>
<td>341 (NNW)</td>
<td>321 (NW)</td>
<td>225 (SW)</td>
</tr>
<tr>
<td>Camera stations</td>
<td>PLM 1 (2681 m a.s.l.)</td>
<td>FIN 1 (2564 m a.s.l.)</td>
<td>RHO 1 (2233 m a.s.l.)</td>
</tr>
<tr>
<td></td>
<td>FIN 2 (3021 m a.s.l.)</td>
<td>RHO 2 (2235 m a.s.l.)</td>
<td>RHO 3 (2392 m a.s.l.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RHO 4 (2589 m a.s.l.)</td>
<td></td>
</tr>
</tbody>
</table>

are lightweight (4 kg for the station plus 4 kg for 8 m of mass balance stakes) and can be mounted by one person.

2.1.2 Camera acquisitions in summer 2019

Figure 3 shows an overview of camera acquisitions and data gaps over the summer 2019.

In total, we obtained 352 daily point mass balance observations between 20 June 2019 and 3 October 2019. The camera longest in the field was on Glacier de la Plaine Morte (91 d between 20 June 2019 and 18 September 2019), while those shortest in the field were two cameras at the tongue of Rhonegletscher (52 d between 13 August 2019 and 2 Octo-
Participants were given the task to read with seven participants. In this kind of experiment all partic-

tor. For nighttime melting (i.e., “(2)”), we equally distribute the overnight melt between the 2 d concerned. This is a trade-off between warmer temperatures before midnight and colder temperatures but longer time span after midnight. For data gaps (point “(3)”), we experienced only short image transmission outages which were mainly due to a 6 d failure in the mobile network connection on Plaine Morte during September 2019. We excluded the daily readings on these days but were able to reconstruct estimates of cumulative mass balance over the time gaps when acquisitions resumed.

2.2 Meteorological input data

To model glacier mass balance, we employ verified products of daily mean and maximum 2 m temperature $T$ and $T_{\text{max}}$, precipitation sum $P$, and mean incoming shortwave radiation $G$ from MeteoSwiss as model input (MeteoSwiss, 2018, 2021a, b). These are delivered on grids with approx. 0.2° spatial resolution, which for Switzerland corresponds to a horizontal resolution of about 2 km.

Temperature uncertainty, given as a root-mean-square error, varies per season from 0.94 K (May to March) to 1.67 K (December to February) in the Alpine region (Frei, 2014; Christoph Frei, personal communication, 2020). We assume a Gaussian-distributed additive error perfectly correlated in time and space for a single glacier but independent on different days. The assumption of a perfect error correlation can be justified by the fact that the station network from which the gridded temperature values are interpolated is much sparser than the scale of individual glaciers. The air temperature lapse rate is derived from the 25 closest cells to a glacier outline centroid using a Bayesian estimation based on a linear regression model:

$$T_{i,t} = e_t + q_t h_i + v_{t,i},$$

where $T_{i,t}$ is the temperature of the $i$th grid cell out of the 25 considered cells at time $t$, $e_t$ is the regression line intercept, $q_t$ is the regression slope (i.e., $\frac{dT}{dz}$), $h_i$ is the height of the $i$th
grid cell, and $v_{t,i} \sim N(0, \sigma^2_{v,t,i})$ are the residuals independent in space and time. Using a g-prior of Zellner (Zellner, 1986), being non-informative in the intercept $c_t$ and model noise variance $\sigma^2_{v,t,i}$ of the regression, we draw samples of the lapse rate $q_t$ from the following posterior distribution:

$$
p(q_t \mid T_t) \propto \left(1 + \left(\frac{q_t - \frac{g}{1+g} \hat{q}_t - \frac{1}{1+g} q_0}{24 c^2}\right)^2\right)^{-25/2}
$$

with

$$
c^2 = \frac{g}{24(1+g)} \sum (h_i - \hat{h})^2 \cdot \left(s_i^2 + \frac{1}{1+g} \sum (h_i - \hat{h})^2(\hat{q}_t - q_0)^2\right).
$$

Above, $p(\cdot)$ means “probability of”, $g$ determines a weighting factor composing the posterior mean (we set $g = 1$), $\hat{q}_t$ is the least squares estimator of the slope, $q_0$ is the prior mean, which we choose to be an annually varying climatological mean gradient at the respective grid location, $\hat{h}$ is the average height of the 25 grid cells, and $s_i^2$ is the residual sum of squares. This is, up to a constant, the probability density of a $t$-distributed random variable with 24 degrees of freedom, shifted by $\frac{(\hat{q}_t + q_0)}{(1+g)}$ and multiplied by $c$. The samples drawn from this distribution are then propagated into the particle filter.

For operational reasons, the precipitation grids contain the 06:00–06:00 local time precipitation sums and thus do not conform with the 00:00–00:00 temperature values (Isotta et al., 2019; MeteoSwiss, 2021b). This might introduce an error, which we cannot quantify. As for temperature, we thus focus again on random errors and pretend for simplicity that the precipitation sum was also from 00:00–00:00. Precipitation uncertainty is generally harder to assess than temperature uncertainty since it involves undercatch errors and skew error distributions. Here, we follow the error assessment by quantiles of precipitation intensity proposed by Isotta et al. (2014) and Christoph Frei, personal communication (2020), who calculated that, for the Alpine region, the standard error at moderate precipitation intensities roughly corresponds to an over- or underestimation by a factor of 1.25. The error increases (decreases) towards low (high) precipitation intensities, and it is generally slightly higher in the summertime. We draw samples from a multiplicative Gaussian error distribution and for the same reason as for temperature, we assume perfect precipitation error correlation at the glacier scale. We also derive Bayesian precipitation lapse rates from the surrounding 25 grid cells in the same fashion as we do for the temperature lapse rate. However, to circumvent high errors in the slope calculation due to the boundedness of precipitation towards zero, we (1) calculate the slope on the square root of the precipitation and (2) assign a probability that the reference has actually received precipitation when the reference cell value is zero but other cells have non-zero precipitation.

Shortwave radiation is derived using data from the geostationary satellite series Meteosat. As an uncertainty, Stöckli (2013) gives a mean absolute bias between 9 and 29 W m$^{-2}$. We assume the errors to be Gaussian and assign a standard deviation of 15 W m$^{-2}$, perfectly correlated on the glacier scale and independent in time. Shortwave radiation is downscaled from the grid to the glacier with potential radiation (see Sect. 3.1).

### 2.3 Glacier outlines and measured mass balances

Glacier outlines for the year 2019 are obtained from GLAMOS (Glacier Monitoring Switzerland), and mass balances are calculated over a fixed glacier surface area (Elßberg et al., 2001; Huss et al., 2012).

For calibration and verification, we use mass balance data acquired in the frame of GLAMOS (Glacier Monitoring Switzerland, 2018). First, intermediate stake readings independent from our near-real-time stations but close to our installations have been made explicitly for this study. Stake locations are depicted in Fig. 1. The reading error for these
measurements is usually estimated to be around 5 cm (e.g., Müller and Kappenberger, 1991). Second, we use seasonal, glacier-wide mass balances based on in situ observations. These observations are acquired during two field campaigns in April and September, respectively. Glacier-wide mass balances are obtained by extrapolating the in situ observations and making the extrapolated values consistent with long-term mass changes. The latter procedure is sometimes referred to as “homogenization” (e.g., Bauder et al., 2007; Huss et al., 2015). For recent years, this homogenization has not been performed since no geodetic mass balances are available yet. The extrapolation method used to infer glacier-wide mass balance from point measurements involves an adjustment of the model parameters of an accumulation and TI melt model (Hock, 1999) at locations where observations are available, while mass balances at grid cells without observations are produced using the calibrated model (Huss et al., 2009, 2015). Uncertainties of the glacier-wide annual mass balance for the measurement period have been estimated to be 0.09–0.2 m w.e. in six experiments in which GLAMOS (1) model parameters (temperature lapse rate, ratio between melt coefficients, summer precipitation correction) and (2) snow extrapolation parameters have been varied within prescribed ranges and for which (3) mass balance stake reading uncertainty, (4) DEM (digital elevation model) and outline uncertainty, (5) climate forcing uncertainty, and (6) point data availability have been accounted.

3 Methods

3.1 Mass balance modeling

Glacier surface mass balance consists of two components: accumulation and ablation. We model accumulation and ablation on elevation bins whose vertical extent is determined by a ∼20 m horizontal spacing of nodes along the central flow line of the glacier (Maussion et al., 2019). To obtain glacier-wide mass balance, node mass balances are weighted with the area per elevation bin. To compute accumulation at different elevations, we employ a simple but widely used accumulation model (e.g., Huss et al., 2008):

\[ c_{sfc}(t, z) = \text{prcp}_{\text{scale}}(t) \cdot P_s(t) \cdot \left[ 1 + (z - z_{\text{ref}}) \cdot \frac{\partial P_s}{\partial z} \right], \]

where \( c_{sfc}(t, z) \) (m w.e.) is the snow accumulation at time step \( t \) and elevation \( z \), \( \text{prcp}_{\text{scale}}(t) \) is the unitless multiplicative precipitation correction factor, \( P_s(t) \) is the sum of solid precipitation at the elevation of the precipitation reference cell \( z_{\text{ref}} \) and time step \( t \), and \( \frac{\partial P_s}{\partial z} \) is the solid precipitation lapse rate. Following Sevruk (1985), we choose \( \text{prcp}_{\text{scale}} \) to vary sinusoidally by ±8% around its mean during 1 year, being highest in winter and lowest in summer. This is to account for systematic variations in gauge undercatch depending on the precipitation phase. The water phase change in the temperature range around 0 °C is modeled using a linear function between 0 and 2 °C; i.e., at 1 °C there is 50% snow and 50% rain (e.g., Maussion et al., 2019).

Since all three glaciers we investigate are in the GLAMOS measurement program and winter mass balance observations are available, the effect of spatial variations in snow accumulation differing from a linear gradient can be incorporated. This is done by choosing a factor \( D(z) \) such that the model mass balance in the elevation bins matches the measured and interpolated distribution of the winter mass balance (Farinotti et al., 2010):

\[ C_{\text{sfc, glamos}}^w(z) = D(z) \cdot C_{\text{sfc}}^w(z). \]  

Here, \( C_{\text{sfc}}^w(z) \) (m w.e.) is the modeled winter surface accumulation, i.e., the sum of individual \( c_{sfc}(t, z) \) over the winter period, and \( C_{\text{sfc, glamos}}^w(z) \) (m w.e.) is the interpolated winter surface accumulation at elevation \( z \).

To model surface ablation, we set up an ensemble of three TI melt models and one simplified energy-balance melt model. We choose this ensemble since the individual models differ in the degree of complexity by which they describe the surface energy balance (Hock, 2003). The models range from using only temperature as input for determining melt via essentially the potential irradiation to using temperature and the actual shortwave radiation. The ensemble contains the following:

1. The “BraithwaiteModel” uses only air temperature as input to calculate melt (Braithwaite and Olesen, 1989; Braithwaite, 1995):

\[ a_{sfc}(t, z) = \text{DDF}_{\text{snow/ice}} \cdot \max(T(t, z) - T_{\text{melt}}, 0), \]

where \( a_{sfc}(t, z) \) (m w.e. d⁻¹) and \( T(t, z) \) (°C) are surface ablation and air temperature at time step \( t \) and height \( z \), respectively. \( \text{DDF}_{\text{snow/ice}} \) (m w.e. K⁻¹ d⁻¹) are the temperature sensitivities (“degree-day factors”) of the surface types (snow and ice), \( \max() \) is the maximum operator, and \( T_{\text{melt}} \) (°C) is the temperature threshold for melt. For this application, we set \( T_{\text{melt}} \) to 0 °C and keep the ratio of \( \text{DDF}_{\text{snow}}/\text{DDF}_{\text{ice}} \) constant at 0.5 (Hock, 2003).

2. The “HockModel” uses potential incoming solar radiation as an additional predictor for melt (Hock, 1999):

\[ a_{sfc}(t, z) = (\text{MF} + a_{\text{snow/ice}} \cdot I_{\text{pot}}(t, z)) \cdot \max(T(t, z) - T_{\text{melt}}, 0), \]

where \( a_{sfc}(t, z) \) (m w.e. d⁻¹) is the temperature melt factor, \( a_{\text{snow/ice}} \) (m w.e. m² d⁻¹ W⁻¹ K⁻¹) are the radiation coefficients for snow and ice, respectively, \( I_{\text{pot}}(t, z) \) (W m⁻²) is the potential clear-sky direct solar radiation at time \( t \) and height \( z \), \( T_{\text{melt}} \) is set again to 0 °C, and the ratio of \( a_{\text{snow}}/a_{\text{ice}} \) is 0.8 (Hock, 1999; Farinotti et al.,
3. The “PellicciottiModel” employs surface albedo and actual incoming shortwave solar radiation (Pellicciotti et al., 2005):

\[
\alpha(t, z) = \left\{ \begin{array}{ll}
TF \cdot T(t, z) + SRF \cdot (1 - \alpha(t, z)) & \text{for } T(t, z) > T_{\text{melt}} \\
G(I_{\text{pot}}(t, z)) & \text{for } T(t, z) \leq T_{\text{melt}} \\
0 & \text{for both}
\end{array} \right.,
\]

where TF (m w \cdot K^{-1} \cdot d^{-1}) is the temperature factor, SRF (m^3 \cdot d^{-1} \cdot W^{-1}) is the shortwave radiation factor, and \(G(I_{\text{pot}}(t, z))\) (W m^{-2}) are the albedo and incoming shortwave radiation at time \(t\) and elevation \(z\), respectively. Note that in this case, \(T_{\text{melt}} = 1^\circ C\) (Pellicciotti et al., 2005).

Albedo is approximated according to the combined decay equation for deep and shallow snow in Brock et al. (2000):

\[
\alpha(t, z) = \left(1 - e^{-\text{swe}(t, z)/\text{swe}^*}\right) \cdot \left(p_1 - p_2 \cdot \log_{10}(T_{\text{acc}}(t, z))\right)
+ e^{-\text{swe}(t, z)/\text{swe}^*} \cdot \left(p_3 - p_4 \cdot T_{\text{acc}}(t, z)\right),
\]

where \(\text{swe}(t, z)\) is the snow water equivalent at time \(t\) and elevation \(z\), \(\text{swe}^* = 0.024\) m w.e. is a scaling length for swe, \(p_1 = 0.713, p_2 = 0.155, p_3 = 0.442,\) and \(p_4 = 0.058\) are empirical coefficients as given in Brock et al. (2000). \(\alpha_u\) is the albedo of the underlying firm/ice below the snow, and \(T_{\text{acc}}(t, z)\) is the accumulated daily maximum temperature > 0 \(^\circ C\) since a snowfall event at elevation \(z\). To avoid infeasible albedo values, \(\alpha(t, z)\) is clipped as suggested in Brock et al. (2000).

4. The “OerlemansModel” calculates melt energy as the residual term of a simplified surface energy balance equation (Oerlemans, 2001):

\[
a_{\text{sfc}}(t, z) = \frac{Q_m(t, z) \delta t}{L_f \rho_w},
\]

where

\[
Q_m(t, z) = (1 - \alpha(t, z)) \cdot G(I_{\text{pot}}(t, z)) + c_0 + c_1 \cdot T(t, z).
\]

Here, \(Q_m(t, z)\) (W m^{-2}) is the melt energy at time \(t\) and elevation \(z\), \(\delta t = 1\) d is a time step, \(L_f = 3.34 \times 10^5\) (J kg^{-1}) is the latent heat of fusion, and \(c_0\) (W m^{-2}) and \(c_1\) (W m^{-2} K^{-1}) are empirical factors. The albedo \(\alpha\) is calculated as well according to Eq. (10).

Figure 4. Calibration workflow used to obtain a prior estimate for the parameters of the model ensemble. \(B_{x, \text{mod}}\) and \(B_{x, \text{glamos}}\) are the modeled and GLAMOS glacier-wide mass balances, respectively, with \(x\) referring either to winter (\(w\)) or annual (\(a\)) values. The yellow arrows highlight the first iteration step, while the green arrows highlight the second iteration step. Figure altered from Huss et al. (2009).

3.2 Mass balance model calibration

For the data assimilation procedure described in Sect. 3.3, we need a prior estimate for the model parameter values of Eqs. (5), (7), (8), (9), and (12). To obtain this, we calibrate the three investigated glaciers against the GLAMOS glacier-wide mass balances (Sect. 2.3) and use an iterative procedure similar to Huss et al. (2009), illustrated in Fig. 4. Additionally, we calibrate the snow redistribution factor \(D(z)\) annually.

Following Huss et al. (2009), all model parameters are initially set to values reported in the literature (Hock, 1999; Oerlemans, 2001; Pellicciotti et al., 2005; Farinotti et al., 2012; Gabbi et al., 2014), and a two-step calibration procedure is then applied: first, the precipitation correction factor is tuned so that the winter mass balance of a given year is reproduced. In this step, the melt factors are held constant at their initial values. In a second step, the calibrated precipitation factor is kept constant, and the melt factors are optimized to reproduce the annual mass balance. The two steps are repeated...
alternately, and both precipitation correction and melt factors converge with every iteration. We terminate the iteration after the absolute difference to the winter and annual mass balance drops below 1 mm w.e.

Once the model parameters have been optimized, we determine the value of $D(z)$ that matches the interpolated winter mass balance. Since this may result in changes in the required model parameters, the iterative procedure is applied one more time as a final step.

3.3 Particle filtering

To ensure that all mass balance model predictions stay within the observational uncertainty, we perform data assimilation. In particular, we employ a particle filter since it does not restrict the class of state transition models and error distributions. Especially when temperatures are around the melting point, the system becomes nonlinear since melt occurs above but not below this point. As a consequence, the distributions we deal with are not necessarily Gaussian. The facts that (a) the temperature chosen to parametrize the melting point is not the same for all four models, (b) the individual model prior distributions are combined to obtain the ensemble prediction, and (c) there can also be accumulation contributing to the overall mass balance add further complexity. We do not use other data assimilation approaches, such as variational methods or ensemble Kalman filtering, because variational methods encounter difficulties when dealing with non-Gaussian priors (van Leeuwen et al., 2019), whilst the ensemble Kalman filter in its original form is not designed for multi-model applications as we use in our case. Overall, particle filtering is a very flexible, generalizable, and readily implementable data assimilation method.

Some extensions of the common particle filter framework allow model parameters and model performance to be estimated over time. With this, we aim at providing optimal, daily mass balance estimates at the glacier scale.

3.3.1 General framework

The general framework for data assimilation consists of a system whose state $x_t$ evolves according to a model, but only partial and uncertain observations $y_t$ of the state are available.

\[
\begin{align*}
x_t &= g(x_{t-1}, \beta_t), & \text{state transition equation} \\
y_t &= \mathcal{H}(x_t) + \epsilon_t, & \text{observation equation} \tag{13}
\end{align*}
\]

Here, $x_{t-1}$ is the state at the previous time step $t-1$, $g(\cdot)$ is the state transition function, $\mathcal{H}$ is the observation operator as introduced in Eq. (1), $\epsilon_t$ is the observation error vector at time $t$, and $\beta_t$ is a random variable that describes model uncertainties. The term for $\beta_t$ does not need to be strictly additive, and it should include uncertainties stemming from the model input variables. The goal of data assimilation is to compute conditional distributions of the system state $x_t$ based on observations $y_{1:t} = (y_1, y_2, \ldots, y_t)$ sequentially for $t = t_0, t_0 + 1, \ldots, t_{\text{end}}$, where $t_0$ and $t_{\text{end}}$ are the time steps with the first and last observations, respectively. In our case, these conditional distributions describe the cumulative mass balance of a glacier, given all available camera observations.

To put this general framework into practice, we use the particle filter, which is a sequential Monte Carlo data assimilation method. Instead of handling conditional distributions of $x_t$ analytically, the particle filter approximates the conditional distribution of a state $x_t$ at time $t$ given the observations $y_{1:t}$ by a weighted sample of size $N_{\text{tot}}$ (e.g., van Leeuwen et al., 2019).

\[
p(x_t | y_{1:t}) \approx \sum_{k=1}^{N_{\text{tot}}} w_{t,k} \delta(x_t - x_{t,k}), \quad \sum_{k=1}^{N_{\text{tot}}} w_{t,k} = 1 \tag{14}
\]

Here $\delta(\cdot)$ is the Dirac delta function, the elements $x_{t,k}$ of the sample are called “particles”, and the weights $w_{t,k}$ associated with the particles $x_{t,k}$ sum to unity.

Usually, particle filtering comprises three repeated steps: the predict step, the update step, and the resampling step. In our case, these steps mean the following: during the predict step, particles holding possible mass balance states are propagated forward in time using the state transition in Eq. (13), where $g(\cdot)$ represents the ensemble prediction of mass balance Eqs. (5)-(10). This acts as a prior estimate of the mass balance distribution. In the update step, the weights of the propagated particles are recalculated based on Bayes’ theorem. This accounts for the information of the next camera observation. In the last step, particles are resampled according to the updated weights. This step is necessary to restore particle diversity that is reduced during the update step. Resampling avoids so-called particle degeneracy, in which all weights collapse on only a few particles. Beyond the common three-step scheme, we additionally estimate model parameters with the particle filter by augmenting the state vector with model parameter values. In this way, we add an additional fourth step to the particle filter scheme, in which we evolve model parameters temporally according to a defined memory parameter. This prevents a collapse of the ensemble due to overconfidence, meaning that model parameter variability would become too low over time.

3.3.2 Application of the framework

The flowchart in Fig. 5 visualizes how the particle filter is implemented in our modeling framework. Figure 6 sheds light on how we perform the individual particle filter steps. The temporal dynamics of the glacier mass balance state can be described by the accumulation model in Eq. (5) combined with the four different melt models in Eqs. (7), (8), (9), and (11). A priori, it is not known which model performs best, and each model has a set of unknown parameters. To take these two uncertainties into account, we augment the state vector by the model index $m_t \in \{1, 2, 3, 4\}$ and
Figure 5. Particle filter workflow during 1 mass budget year ("MB year"). We use uncertain model estimates to predict mass balance with 10 000 particles and reset the cumulative mass balance when a camera is set up. The model mass balance estimate is updated at time steps for which observations are available. To avoid overconfidence of the particle filter, we apply a partial resampling technique. The individual particle filter steps are sketched in Fig. 6.

Figure 6. Illustration of the individual particle filter steps. The example refers to a case in which four models (blue, orange, red, and green) start with two particles each. The blue curve represents the observation distribution. At time step $t_0 + 1$, the green model performs poorly and receives entirely low weights during the update step (weights are shown by the size of the circles). In the resampling step, we modify the weights of the other particles again. This is for not omitting the green model entirely due to temporarily poor performance. As the green model stays in the ensemble, it can recover later, i.e., when making a good prediction (here: $t_0 + 2$).
the model parameters $\theta_i$. In this way, a model weight and the model parameter values are also estimated based on the observations. Although the unknown parameters are different for each model, we do not use an additional model index for $\theta_i$. Instead, we ensure that for all particles, $\theta_{t,k}$ is always the parameter vector associated with model $m_{t,k}$. This means that, when following a given particle backwards in time, its entire dynamics is governed by one single model only. In the forward direction, a particle can “die off” and be replaced by another particle with different model index. In this case, both the model index $m_{t,k}$ and the entire past trajectory are changed to the new model.

As the state has to provide all information that is needed to predict the next observation, we also include the surface albedo and the snow water equivalent on the ice in our state vector. Hence, the state vector is defined as

$$x_t = (m_t, \theta_t, \xi_t), \quad \xi_t = (b_{\text{sfc}}(t,z), \alpha(t,z), \text{swe}(t,z)).$$

Here, $\xi_t$ is called the physical state.

### 3.3.3 Predict step

During the predict step, the explicit temporal evolution of the physical state $\xi_t$ involves the randomized error accounting for uncertainties in the meteorological input (Sect. 2.2). Here, we call these errors $\eta_t$ and set an additional scalar subscript to indicate that the errors are different for each meteorological variable. We first predict $c_{\text{sfc}}(t,z)$, $T_{\text{acc}}(t,z)$, and $\text{swe}(t,z)$.

$$c_{\text{sfc}}(t,z) = c_{\text{sfc}}(P_{\text{sfc}}(t,z), \eta_{t,2}, \theta_t)$$

$$T_{\text{acc}}(t,z) = \begin{cases} T_{\text{acc}}(\alpha(t-1,z)) + T_{\max}(t,z) + \eta_{t,1}, & \text{if } T_{\max}(t,z) > 0 \\ c_{\text{sfc}}(t,z) < 0.001 \text{ m w.e. d}^{-1} \\ 0, \text{ otherwise}. \end{cases}$$

$$\text{swe}(t,z) = \max(\text{swe}(t-1,z) - a_{\text{sfc}}(t-1,z), 0) + c_{\text{sfc}}(t,z)$$

Based on Eqs. (7), (8), (9), and (11), the predicted mass balance is then

$$b_{\text{sfc}}(t,z) = b_{\text{sfc}}(t-1,z) + c_{\text{sfc}}(t,z) - a_{\text{sfc}}(T(t,z) + \eta_{t,3}, G(I_{\text{pot}}, t, z) + \eta_{t,4}, \alpha(T_{\text{acc}}(t,z)), \text{swe}(t,z), m_t, \theta_t) + \beta_t,$$

where the errors $\eta_t$ are independent in time but partly perfectly correlated in space for reasons described in Sect. 2.2. By introducing both the meteorological uncertainty $\eta_t$ and the parameter uncertainties, we shift the majority of the uncertainty contained in $\beta_t$ to these variables. Since the remaining uncertainty for $\beta_t$ is small and hard to quantify, we set $\beta_t = 0$ for simplicity. With this assumption, we neglect some additional uncertainty contained in $\beta_t$, being aware that this might lead to “jumps” in the temporal evolution of the model performance. Finally, the observations $y_t$ depend only on the cumulative mass balance at the elevation $z$ of the camera as specified in Eq. (1).

We use a total of $N_{\text{tot}} = 10000$ particles and set the weights at time $t_0$ (i.e., the time when the first camera observation is available) to $1/N_{\text{tot}}$. Since at $t_0$ all models have equal probabilities, $N_{\text{tot}}/4$ particles are assigned to each of the four models. The initial value of $b_{\text{sfc}}(t_0, z)$ is set to zero for all particles, whereas $\alpha(t_0, z)$ is determined by the maximum air temperature since the last snowfall before $t_0$, and $\text{swe}(t_0, z)$ depends on the cumulative mass balance before $t_0$. Finally, the initial calibration parameter values $\theta_{t_0,k}$ of the particles with model index $j$ are obtained by drawing Monte Carlo samples from a normal distribution fitted to the logarithmized parameters of model $j$, as they were calibrated in the past (see Sect. 3.2). Table 2 shows the means and standard deviations for the input parameters of the three glaciers.

### 3.3.4 Update step

In the update step, all particles are then reweighted by multiplying the density of the observations $y_t$ given the state of individual particles $x_{t,k}$ with their respective weights at $t-1$ and normalizing the weights to sum to unity (van Leeuwen et al., 2019).

$$w_{t,k} = w_{t-1,k} \frac{p(y_t | x_{t,k})}{\sum_j w_{t-1,j} p(y_t | x_{t,j})}$$

In our case, $y_t = h(t,z)$, and $p(y_t | x_{t,k})$ is the normal density with mean $b_{\text{sfc}}(t,z)k/\rho_{\text{bulk}}$ and standard deviation $\sigma_e$, evaluated at $h(t,z)$. After updating the model predictions with the observations, we are interested in (a) the posterior model probabilities $\pi_{t,j}$, (b) the posterior distribution of model parameters $\theta$, and of course (c) the posterior distribution of the physical state given all observations $y_{1,t}$. These quantities can be decomposed from the approximation with weighted particles in Eq. (14). The posterior model probability is given by

$$p(m_t = j | y_{1,t}) \approx \pi_{t,j} = \sum_{k=1}^{N_{\text{tot}}} w_{t,k} \delta(m_{t,k} - j),$$

where $\pi_{t,j}$ is the approximation of the posterior model probability at time $t$ and model $j$. The posterior distribution of the parameters of model $j$ is approximated by

$$p(\theta_t | y_{1,t}, m_t = j) \approx \sum_{k=1}^{N_{\text{tot}}} \frac{w_{t,k}}{\pi_{t,j}} \delta(m_{t,k} - j) \delta(\theta_{t,k} - \theta_t).$$

The posterior distribution of the physical state takes the model uncertainty into account. It combines the posterior distributions under the different models $j$ according to the law
of total probability, in which we can insert Eqs. (21) and (22):

\[
p(\xi_t | y_{1:t}) = \sum_{j=1}^{4} p(m_t = j | y_{1:t}) p(\xi_t | y_{1:t}, m_t = j) \\
\approx \sum_{j=1}^{4} \pi_{1,j} \sum_{k=1}^{N_{\text{tot}}} w_{t,k} \delta(m_t = j - j) \delta(\xi_t,k - \xi_t) \\
= \sum_{k=1}^{N_{\text{tot}}} w_{t,k} \delta(\xi_t,k - \xi_t). \quad (23)
\]

As the observations only measure the mass change since the installation of a camera, a difficulty occurs if several cameras are installed on different days at different elevations of the same glacier. We elaborate on the technical details for these cases in Appendix A.

### 3.3.5 Resampling

During the resampling step, the updated weights are used to choose a new set of \(N_{\text{tot}}\) particles with equal weights. To achieve equal weights, particles with low weights are removed, whereas those with high weights are duplicated. Because there is no stochasticity in the evolution of \(m_t\), though, when the particle index \(k\) is fixed, for some models only a few particles with the according model index survive after a couple of iterations. If this occurs, the respective model has little chance to become better represented at later time steps, which is unfavorable, since the model might give better predictions on average.

To overcome this problem, we assign a minimum contribution to each model of the ensemble, regardless of the model’s performance at a certain time step. To compensate for the potentially too high resampling rate of a poor prediction, we lower the weights of all particles of a model whose contribution has been deliberately increased to match the chosen minimum contribution. In turn, we increase the weights of all other particles to compensate for their underrepresentation. This means that on average, the original weights per model remain unchanged. For technical details of the resampling procedure, see Appendix B.

### 3.3.6 Parameter evolution

The dynamics of the augmented state is defined such that the model index does not change over time. However, parameters are evolved temporally such that after a long period without observations, \(\theta\) is distributed according to the prior parameter distribution.

\[
\theta_{t+1} = \rho \theta_t + (1 - \rho) \mu_0 + \xi_t, \quad \xi_t \sim \mathcal{N}(0, (1 - \rho^2) \Sigma_0) \quad (24)
\]

Here, \(\mu_0\) and \(\Sigma_0\) are the prior mean and the prior covariance of \(\theta\) at the starting time \(t_0\), which we determine from the calibration procedure in Sect. 3.2, and \(\rho \in [0; 1]\) is a memory parameter that we choose to be 0.9. This step accounts for the fact that parameters are not necessarily constant in time, and it also ensures the reintroduction of parameter diversity which is lost during the resampling step.

### 3.4 Validation scores

To validate the daily mass balance predictions made with the particle filter, we use the CRPS (continuous ranked probability score). The CRPS is designed to estimate the deviation of a probabilistic forecast from an observation. It takes into account both the deviation of the median forecast from the actual observation (forecast reliability) and the spread of the forecast distribution (forecast resolution). This means that a forecast close to the observation median can still receive a poor CRPS if the forecast distribution spread is high, and the other way around. Lower values of the CRPS correspond to better forecasts. The minimum value is zero, corresponding to a perfect, deterministic forecast of the observation.

The CRPS is defined as (Hersbach, 2000)

\[
\text{CRPS} = \int_{-\infty}^{\infty} \left[ P_t(b_{sfc}/\rho_{\text{bulk}} \cdot \rho_w) \\
- H(b_{sfc}/\rho_{\text{bulk}} \cdot \rho_w - h(t, z))^2 \right] db_{sfc}, \quad (25)
\]

where \(P_t(\cdot)\) is the forecast mass balance cumulative probability distribution, and \(H(\cdot)\) is the Heaviside function. The usual
choice for $P_1$ is the weighted ensemble distribution of the particles from the predict step, i.e., a discrete step function with jumps of height $w_{t-1,k}$ at the positions $H(b_{fc}(t,z)k)$, where $b_{fc}(t,z)k$ are the prediction particles. Note that this setting does not account for the observation error of $h(t,z)$, implying that the score is not “proper”; i.e., it does not always return the best value when the prediction distribution is the true distribution (Ferro, 2017; Brehmer and Gneiting, 2019). To obtain a proper score, one can use the forecast of the camera reading $h(t,z)$, which is the Gaussian mixture with weights $w_{t-1,k}$, mean values $H(b_{fc}(t,z)k)$, and common variance $\sigma_e^2$. Despite being proper, it still has some theoretical shortcomings (Ferro, 2017). Since for our data the values of the two scores do not differ much, we use only the proper score in all result figures but give also the value of the common CRPS in square brackets in the text.

4 Results and discussion

4.1 Mass balance observations

Figure 7 shows the observed cumulative mass balance at the individual cameras, an example of meteorological conditions at station FIN 1 (providing the longest time series), daily mass balance rates at FIN 1, and four example camera images.

Considering all stations, we have observed ice melt rates of up to 0.12 m.w.e.d$^{-1}$ and a cumulative mass balance of about $-5.5$ m.w.e. in 81 d close to the terminus of Findelgletscher (FIN 1). Different camera stations reveal different melt rates and total ablation, which generally depend on the station’s elevation. However, stations at different elevations can have similar melt rates as well. For example, station PLM 1 in 2589 m.a.s.l. experienced an average melt rate of $-0.047$ m.w.e.d$^{-1}$, while the average melt rate at FIN 2 (3015 m.a.s.l.) was $-0.043$ m.w.e.d$^{-1}$ during the same period (we count only days with net ablation). Further, station PLM 1 had the lowest average melt rate despite not being at the highest elevation. This might be due to the meteorological conditions, such as the formation of local cold air pools, and the so-called “Massenerhebung effect” (Barry, 1992). The latter describes the tendency of higher temperatures to occur at the same elevation in the inner Alps as on their outer margins. For all stations, average melt rates during July and August are similar (0.073 ± 0.012 m.w.e.d$^{-1}$ in July vs. 0.062 ± 0.011 m.w.e.d$^{-1}$ in August) and 0.02–0.03 m.w.e.d$^{-1}$ lower (i.e., 0.044 ± 0.014 m.w.e.d$^{-1}$) in September. On Glacier de la Plaine Morte, the difference is most pronounced with a drop of 0.06 m.w.e.d$^{-1}$ between August and September. Again, this is probably caused by local effects. On average, the difference between minimum and maximum melt measured at different stations on a particular day was 0.035 m.w.e.d$^{-1}$. Over the observational period, this difference ranged from 0.005 to 0.081 m.w.e.d$^{-1}$. The highest difference (0.081 m.w.e.d$^{-1}$) occurred on 1 September 2019 in connection with the passage of a convergence line and cold front (German Meteorological Service, 2019): while Glacier de la Plaine Morte was already under the influence of cooler weather, Findelgletscher and Rhonegletscher experienced another melt-intensive day. The variability at individual stations, measured as standard deviation of a 14 d running mean, was generally low during July and August (0.016 m.w.e.d$^{-1}$) and increased at the beginning of September (0.026 m.w.e.d$^{-1}$). We attribute this increase to the onset of intermittent snowfalls at individual sites.

As shown by the pictures from station FIN 1 (Fig. 7d), summer 2019 is characterized by a variety of events, ranging from very hot, melt-intensive days to snowfalls at high elevations. The time series of normalized mean daily temperature and shortwave radiation at station FIN 1 (Fig. 7b) illustrate that two heat waves occurred at the end of June and end of July 2019. The total amount of water released by snowmelt and ice melt on Swiss glaciers during these heat waves was 0.8 km$^3$ (Swiss Academy of Sciences, 2019), which approximately equals the annual amount of drinking water consumed in the country. These extreme phases are also mirrored in the melt observed at our stations (Fig. 7): for FIN 1, daily melt rates peaked between 0.09 and 0.12 m.w.e.d$^{-1}$. For days with a range-normalized temperature exceeding 0.8 (9 d in total, Fig. 7b), the average melt rate is 0.1 m.w.e.d$^{-1}$ at that station. During these nine days, modeled, glacier-wide melt indicates the release of $6 \times 10^6$ m$^3$ of water. Another phase with very high melt occurred at the end of August 2019. Here, normalized temperature and radiation are average (mean values of 0.6 and 0.5, respectively). The exact causes for this strong melt event are unclear, and we speculate that it might be related (at least in part) to rain events that were not captured by the meteorological input despite being visible in our camera images between 28 and 31 August 2019. Summer melt phases were also interrupted by two snowfalls of different strengths: small amounts from 15 July 2019 (image 2 of Fig. 7d) and larger amounts, summing up to 0.25 m snow height in total, in early September (image 4 of Fig. 7).

4.2 Particle filter mass balance validation

Besides the direct observations presented above (Sect. 4.1), our framework enables predictions of daily mass balance. In this section, these predictions are (i) validated against reference forecasts (Sect. 4.2.1), (ii) cross-validated against test subsets of observations (Sect. 4.2.2), and (iii) compared against glacier-wide mass balances reported by GLAMOS (Sect. 4.2.3).
Validation against reference forecasts

We consider two types of reference forecasts: first, a forecast with (i) mean glacier-wide melt parameters as obtained from past calibration (Sect. 3.2) and (ii) the precipitation correction factor $prcp_{scale}$ constrained by the 2019 GLAMOS winter mass balance; and second, a forecast with a partially informed model including the same constraint for $prcp_{scale}$ but also a tuning of the melt parameters to reproduce one further intermediate point measurement. The latter measurement is the cumulative mass balance between September 2018 and 2019 at the mass balance stake closest to each camera (locations in Fig. 1). Since there are up to four stake readings per glacier, we calculate single parameter sets tuned to reproduce all possible combinations of stake readings per glacier. This results in 19 CRPS values in total, for which we calculate the median. We also distinguish between the case in which the uncertainties in the meteorological inputs are taken into account and the case in which they are not.

Finally, we calculate the CRPS for the two reference forecasts by inserting two different values into the CRPS equation: (a) the mass balance of each day separately and (b) the cumulative mass balance. Note that for the particle filter, there is no need to make this distinction. Indeed, the daily deviation from a mass balance observation also equals the deviation from the cumulative observation. Figure 8 shows the results of the validation.

For the particle filter, daily and cumulative melt observations are generally reproduced well, with an average CRPS of 0.012 [0.012] m (proper CRPS outside and standard CRPS inside the square brackets). At the end of the assimilation period, Rhonegletscher has an average CRPS of 0.017 m, which is almost double the CRPS for Findelgletscher (CRPS = 0.01 m) and Glacier de la Plaine Morte (CRPS = 0.008 m). The high value of Rhonegletscher is related to the switch-on of cameras RHO 1 and RHO 3. Indeed, the glacier also has CRPS $\approx$ 0.01 m before that. Poor predictive performances also occur after snowfalls, probably related to the uncertainties by which the mass balance stake can be read during these times. We run experiments in which the particle filter is limited to using mean parameters and/or single models instead of parameter distributions and the model ensemble. In the experiments, the resulting average CRPS values are higher than the average CRPS obtained with the ensemble and time-variant parameters. The lowest single values occur for specific combinations when running the particle filter with the BraithwaiteModel and Oerlemans-Model and flexible parameters on Glacier de la Plaine Morte.
Figure 8. Median CRPS values over “n” validation cases for different forecasts. The following symbols are used: $\hat{\theta}$ = mean parameters from past calibration; $\theta_{\text{obs}}$ = parameters calibrated on different combinations of mass balance stake observations close to the cameras; $\theta_{\text{pf}}$ = parameters found with the particle filter. Cases accounting for (red dots and analyses with “$\eta$”) and neglecting (blue dots) the uncer-
tainty in the meteorological variables are distinguished, and “cum” and “day” indicate the errors in the cumulative and daily mass balance predictions, respectively. For the particle filter (highlighted in green), the label “cum/day” indicates that the two errors coincide, and ‘j’ indicates cases where the particle filter was run with only one model.

Note that if no probabilistic temperature and precipitation lapse rate is used, the resulting CRPS values from the experiments with mean parameters and/or only one model are even higher than the highest CRPS obtained using the ensemble and time-variant parameters. The experiments thus show that it is beneficial to include all four models and parameter uncertainty into the particle filter.

Comparing the CRPS of the particle filter with the reference forecasts, the performance closest to the particle filter is delivered by the forecast produced with mean melt parameters and no uncertainty in the meteorological input (mean CRPS = 0.013 [0.015] m). When the CRPS is calculated from the cumulative mass balance produced with mean melt parameters, the CRPS increases to 0.333 [0.243] m on average. This is because the mean parameters do not adapt to the meteorological conditions over time, and in this case, the cumulative mass balance can temporally be under- or overestimated or even diverge completely over time. Somewhat counterintuitively but for the same reason, the CRPS is similar when parameters have been tuned to match nearby stake readings. For the cumulative deviation, we find CRPS = 0.25 [0.251] m w.e. when considering meteorological uncertainty, and CRPS = 0.294 [0.28] m w.e. when not doing so. Compared to both the particle filter prediction and the prediction with mean melt parameters, the CRPS of daily mass balances produced without considering meteorological uncertainty is slightly higher (median CRPS: 0.023 [0.025] m w.e.).

In general and for the individual glaciers, the particle filter improves the CRPS of the reference forecasts by 95% to 96%. For the daily forecasts, the performance of the particle filter is only partly better, with improvements in CRPS between 8% and 48%. Along with the performance, a further important advantage of the particle filter is that it provides daily estimates for the results’ uncertainties without need for further calculations. Indeed, this information can be essential, especially for the operational application of our framework.

4.2.2 Cross-validation

A different approach for validating the particle filter is to only use subsets of the available camera observations as input and to evaluate the predicted mass balances at the remaining locations (cross-validation). We do so by splitting the available observations into training and test subsets of cameras, i.e., by keeping the time series of a given station together (as opposed to splitting individual time series). Our test sets always contain one time series; i.e., we perform a leave-one-out cross-validation. Figure 9 shows the temporal evolution of the CRPS at the test locations, i.e., at the stations not used by the particle filter.

We find that, in general, the cumulative mass balance at the test locations follows the cumulative observation curve well but not as closely as when the test location’s data are assimilated with the particle filter. This shows the benefit of having several cameras per glacier mounted at different elevations. For Findelgletscher, we find 8.8% average deviation (median CRPS of 0.071 and 0.108 m w.e. for the two stations) when comparing the cumulative mass balance curve with the particle filter prediction. For Rhonegletscher, the average deviation at the test locations is 9.0% average deviation (median CRPS 0.14, 0.148, 0.067, and 0.178 m w.e. depending on the station). The highest CRPS values for Rhone stem from the period after the middle of August when two additional cameras were set up, but the values still outperform the reference forecasts of Sect. 4.2.1.
The temporal pattern evident in Fig. 9 includes an increasing CRPS through time but at different rates depending on the cross-validation subset. The individual pattern originates from (1) a stations' representativity for the given elevation band it is located in, (2) the combination of stations in the cross-validation subsets, and (3) cumulative error characteristics, since we observe cumulative mass balance over time. Station RHO 3, for example, can generally be modeled with lower errors compared to other stations. We speculate this being related to its location, which is in a relatively flat area with little crevasses. The other stations are instead either in the vicinity of crevasses (RHO 4) or influenced by shadows from the surrounding terrain, dark glacier surface, or steep ice (RHO 1 and RHO 2). RHO 1 and RHO 2 also show that even neighboring stations can exhibit different melt. This affects the results of the cross-validation whenever one of these two stations is excluded from the training dataset.

The above results show the ability of the particle filter to also predict melt at locations without observations, albeit with a lower performance when compared to the situation in which all observations are assimilated. The results also show that even with an augmented particle filter, it is demanding to find a unique, glacier-wide parameter set that correctly reproduces the mass balance at all locations.

### 4.2.3 Comparison to GLAMOS glacier-wide mass balances

We compare our assimilated model ensemble predictions to the glacier-wide annual mass balance reported by GLAMOS at the autumn field date of the mass budget year 2019. We do so by running the model from the field campaign date in autumn 2018. Figure 10 illustrates the different model and parameter settings used during the simulation.

During the period preceding the installation of our cameras, we calculate mass balance with the parameters calibrated in Sect. 3.2. This results in about 45 distinct model runs, which we call “free model runs”. We use this first period to provide initial conditions for the particle filter period, which lasts from the first camera setup on a respective glacier either until cameras are retrieved or until the autumn field date is reached (whatever comes first). To achieve a connection between the free model run and the period during which the particle filter is used, we sample 10 000 times from the initial conditions at the first camera setup date. We refer to this procedure as “particle filtering without pre-selection (of initial conditions)”. Not all free model runs have to be used, though: they can also be pre-selected based on the cumulative mass balance observed at the stakes closest to the camera stations. For this case, we select model runs that reproduce these observations within an estimated reading uncertainty of ± 0.05 m w.e. (“particle filtering with pre-selection”). The cumulative mass balances calculated with these two procedures are compared to the GLAMOS analyses in Table 3.

For particle filtering without pre-selection of initial conditions, the difference from the GLAMOS analyses is 0.67 m w.e. for Rhonegletscher, 0.2 m w.e. for Findelgletscher, and 0.05 m w.e. for Plaine Morte. With pre-

### Table 3. Mass balances calculated with the particle filter between the autumn field dates of 2018 and 2019 against the values reported by GLAMOS. See text for the difference of particle filtering with and without pre-selection. Uncertainty values are given as standard deviations.

<table>
<thead>
<tr>
<th>Glacier</th>
<th>Particle filter (no pre-selection) (m w.e.)</th>
<th>Particle filter (pre-selection) (m w.e.)</th>
<th>GLAMOS (m w.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLM</td>
<td>-1.74 ± 0.29</td>
<td>-1.79 ± 0.38</td>
<td>-1.77 ± 0.09</td>
</tr>
<tr>
<td>FIN</td>
<td>-0.04 ± 0.76</td>
<td>-0.48 ± 0.27</td>
<td>-0.24 ± 0.16</td>
</tr>
<tr>
<td>RHO</td>
<td>-0.09 ± 0.90</td>
<td>-0.84 ± 0.28</td>
<td>-0.77 ± 0.20</td>
</tr>
</tbody>
</table>

Figure 9. Temporal evolution of the CRPS as determined in a leave-one-out cross-validation procedure on Rhonegletscher and Findelgletscher. “TRAIN” and “TEST” stand for the stations assimilated by the particle filter and the station used for the validation, respectively.
Figure 10. Schematic model and parameter settings on Rhonegletscher during the mass budget year 2019. After an initial phase with parameters from past calibration, the precipitation correction factor \( \text{prcp}_{\text{scale}} \) is tuned to match the winter mass balance. When the first camera is set up, we sample the existing model runs 10,000 times to be able to couple the free model runs with the 10,000 particles during the particle filter period (not all are drawn for readability).

4.3 Individual model performance

We analyze model performance by considering the temporal evolution of the model probabilities \( \pi_{t,j} \) and model particle numbers \( N_{t,j} \) for the four melt models over time at individual glaciers. High model performance is indicated by high probabilities and large particle numbers over long time periods.

Figure 11 shows the model performance of all four melt models and at all three glaciers.

In general, we find that the model probabilities are sensitive to the ensemble input, such as the parameter priors, and the prescribed meteorological uncertainty. This is an indication of the ensemble choosing the model combination that best reproduces the observations at any time. Note that none of the models is removed from the ensemble in the resampling step even when the model performs poorly. During the assimilation period, indeed, models can recover and can show good performances at a later stage (see, for example, the HockModel for Rhonegletscher or the PellicciottiModel for Findelgletscher). This shows the utility of the resampling procedure introduced in Sect. 3.3.5.

During the assimilation period of an individual glacier, often one model dominates the ensemble for a given amount of time (“model dominance” being the case in which the model probability is > 0.5). Model dominance, and especially fast switches between dominant models, can be indicative of a mode collapse, resulting from either an overconfident likelihood or prior, or both, operating in an M-open framework (Bernardo and Smith, 2009), i.e., the case in which the “true”
model is not a choice amongst the available models. In our case, we believe that the ensemble prior might be overconfident on average since we have chosen the observational error conservatively; i.e., we have chosen the largest errors emerging in the round-robin experiment (Sect. 2.1.2). This would lead to a model preferably obtaining high weights, which had already dominated on the previous days. However, when the likelihood is overconfident or there is strong evidence that a previously well-performing model now performs worse, the filter might switch back and forth between individual models that best describe the observations. We accept this model dominance and the fast switching as a sign that the overall ensemble performance is improved. Averaged over all glacier and time steps, the PellicciottiModel has the highest model probability (0.39), while the BraithwaiteModel has an average model probability of 0.24, the OerlemansModel of 0.23, and the HockModel of 0.14. The relatively high probabilities assigned to the PellicciottiModel can have various reasons, and we suspect that two are of particular importance in our case: first, the calibration might have led to a broad prior parameter distribution, allowing for the model to adapt to various combinations of meteorological input and observed melt. Second, using the actual solar radiation $G$ instead of the potential irradiation $I_{pot}$ might provide a further advantage since this accounts for partly cloudy conditions and diffuse radiation, which the potential irradiation is not able to cover. The fact that the second highest probability is assigned to the OerlemansModel can have various reasons, and we suspect that two are of particular importance in our case: first, the calibration might have led to a broad prior parameter distribution, allowing for the model to adapt to various combinations of meteorological input and observed melt. Second, using the actual solar radiation $G$ instead of the potential irradiation $I_{pot}$ might provide a further advantage since this accounts for partly cloudy conditions and diffuse radiation, which the potential irradiation is not able to cover.

Figure 13 shows the evolution of the distribution of individual model parameters during the assimilation period. The example refers to Findelgletscher.

Three phases of quick parameter changes can be observed. First, the parameters change rapidly on the first days of the assimilation period. This means that the prior parameter distributions do not match the exact parameter distributions needed to model the mass balance at the camera locations. This is due to both the calibration time span (seasonal calibration vs. daily application) and the low sample size of the calibrated parameters. A second rapid change can be observed after the second camera has been switched on, i.e., on 24 July 2019. Here, an adjustment in the parameters is needed in order to accommodate the mass balance at both stations equally well. The third rapid change starts when ablation at station FIN 1 is highest but when radiation and temperature are not at their maximum. Here, the change might be due to the model being forced to yield high ablation rates despite only moderate meteorological forcing. This shows the
advantage of employing the model ensemble as opposed to, for example, a single model with deterministic parameters: the ensemble also reproduces system states which cannot be explained by the uncertain meteorological input.

5 Conclusions

In this study, we mounted seven cameras on three Swiss glaciers, delivering 352 point mass balance observations throughout summer 2019. At the camera locations, we observed daily melt rates of up to 0.12 m w.e. d$^{-1}$ and cumulative melt of up to $\sim$ 5 m w.e. in 81 d. To calculate near-real-time mass balances, we used an ensemble of three TI mod-
els, a simplified energy balance model, and meteorological 
model input. The camera observations were assimilated into 
the model ensemble by using a specifically developed parti-
cle filtering scheme. The particular focus was put on deliver-
ing a stable ensemble capable of reproducing glacier mass 
balance throughout the summer. Variability in the model 
parameters, as well as particle filter stability, was consid-
ered. For the former, a prior parameter distribution obtained 
from calibration against past seasonal glaciological mass bal-
ances was used as input to an augmented particle filter capa-
cible of estimating parameters while assimilating observa-
tions. For the latter, the particle filter was designed such 
that models with temporarily poor performance can recover at 
a later stage. For the mass budget year 2019, we calculate 
cumulative mass balances of \(-1.79\) m.w.e., \(-0.48\) m.w.e., 
and \(-0.84\) m.w.e. for Glacier de la Plaine Morte, Findelgletscher, 
and Rhonegletscher, respectively.

The mass balances given by the particle filter were closer 
to the cumulative observations (CRPS = 0.012 m.w.e.) than 
two reference forecasts which either assumed no measure-
ments to be available or only used one intermediate set of 
state readings. Measured with the CRPS for cumulative mass 
balances, the particle filter improves the performance of re-
ference forecasts by 95 % to 96 %. As a further advantage, the 
particle filter delivers direct uncertainty estimates. A leave-
one-out cross-validation procedure showed that the cumula-
tive mass balance predicted with the particle filter is within 
9 % of the observations at any location. In an analysis of the 
individual model performance, we found that our technique 
to prevent models from being removed from the ensemble is 
useful since models can recover at a later stage. In terms of 
model ensemble, the TI model by Pellicciotti et al. (2005) 
achieved the highest average model probability (0.39). None 
of the four models has an average probability \(< 10 \%\), and 
even if individual models can temporarily perform poorly, 
our technique preventing models from being removed from 
the ensemble completely allows them to recover at a later 
stage. Fast temporal switches between model probabilities 
are attributed to overconfident likelihood or prior distribu-
tions, or both. As a future venture, we envision an extensi-
on of the particle filter in which glacier mass balances and model 
parameters are further constrained by remotely sensed ob-
vervations of albedo and snow lines. These measurements are 
indirect but have the potential to (1) complement the cam-
era observations extensively and to (2) overcome the limited 
knowledge about the spatial and temporal extrapolation of 
glacier mass balances and model parameters.

Appendix A: Handling of multiple cameras

Assume that camera \(i\) is installed at elevation \(z_i\) on day \(t_{i-1}\), 
where \(t_0 < t_1 < t_2 \ldots\) (to be coherent with earlier notation 
that the first camera is installed at time \(t_0\)). From time \(t_{i-1}\) 
 onwards, we include \(b_{sfc}(t_{i-1}, z_i)\) in the state vector as a 
component which remains constant. Then the observations 
at time \(t > t_{i-1}\) are functions of the state at time \(t\):

\[
h(t, z_i) = \frac{b_{sfc}(t, z_i) - b_{sfc}(t_{i-1}, z_i)}{\rho_{bulk}} + \epsilon(t, z_i). \tag{A1}
\]

The true value of \(b_{sfc}(t_{i-1}, z_i)\) is unknown, and the uncer-
tainty is represented by the values \(b_{sfc,k}(t_{i-1}, z_i)\) of the 
particles. Thus at time \(t\), the contribution from the observation 
h(t, z_i) to the weight of particle \(k\) is proportional to

\[
\exp\left(-\frac{(h(t, z_i) - (b_{sfc,k}(t_{i-1}, z_i))/\rho_{bulk} - \rho_w)^2}{2\sigma^2}\right). \tag{A2}
\]

Although \(b_{sfc,k}(t_{i-1}, z_i)\) never changes during the propaga-
tion step, it will change in the resampling steps. Thus the uncer-
tainty about \(b_{sfc}(t_{i-1}, z_i)\) will decrease as time proceeds. 
This is presumably not realistic, but the effect of small errors 
in the baseline also diminishes as time proceeds.

Appendix B: Resampling procedure

The technical details of the resampling procedure in 
Sect. 3.3.5 are the following: if, after prediction and update, 
\(N_{t,j}\) denotes the number of particles with model index \(j\), 
we prevent models from not being resampled by choosing 
a minimum model contribution \(\phi < \frac{1}{4}\) to the ensemble. 
This ensures that the resampling step preserves a minimum parti-
cle number \(N_{t,j} \geq \phi N_{tot}\) representing model \(j\). For our 
application, we choose \(\phi = 0.1\). If the posterior probability 
of model \(j\) (Eq. 21) is smaller than the minimum contribution 
\(\phi\), an unweighted sample that represents \(\pi_{t,j}\) correctly must 
have less than \(\phi N_{tot}\) particles with model index \(j\). To ensure 
our minimum contribution condition though, we generate a 
weighted sample \((\tilde{x}_{t,k}, \tilde{\omega}_{t,k})\) such that each model index \(j\) 
appears at least \(\phi N_{tot}\) times, and the weights are as close to 
uniform as possible. We select the particles \(\tilde{x}_{t,k}\) in a two-
step resampling procedure: first, the number \(N_{t,j}\) of parti-
cles with model index \(j\) is chosen to be \(N_{t,j} = \phi N_{tot} + L_{t,j}\), 
where \(L_{t,j}\) are excess frequencies. We obtain these frequen-
cies by sampling a total of \(N_{tot}(1 - 4\phi)\) model indices from 
\([1, 2, 3, 4]\) with weights proportional to how much a model 
probability exceeds the chosen minimum contribution, i.e., 
\(\max(0, \pi_{t,j} - \phi)\). In a second step, we draw for each model 
a resample of size \(N_{t,j}\) with weights \(w_{t,k}/\pi_{t,j}\) from the 
particles with model index \(j\). The combined set of the \(N_{tot}\) re-
sampled particles gives the new filter particles \(\tilde{x}_{t,k}\).

However, introducing a restriction on the minimum num-
ber of particles per model can lead to biased estimates as poor 
models with probability \(\pi_{t,j} \leq \phi\) are overrepresented in 
the ensemble. To compensate for poor models occurring too of-
ten among the resampled particles (and the other models not 
onenough), the following weight has to be given to \(\tilde{x}_{t,k}\):

\[
\tilde{\omega}_{t,k} = \frac{\pi_{t,j}}{N_{t,j}} \text{ if } \tilde{m}_{t,k} = j. \tag{B1}
\]
These weights sum to unity and preserve the original weights $w_{t,k}$ on average. Since they can become very small though, we work with the logarithm of the weights to avoid numerical underflow. It should be noted that we insert $\tilde{w}_{t-1,k}$ for $w_{t-1,k}$ in Eqs. (14) and (20). In order to see that the weights we choose for $\tilde{x}_{t,k}$ are correct, we denote the number of times the particle $x_{t,k}$ is selected in the resampling procedure by $\tilde{M}_{t,k}$. This means that the resampling gives $x_{t,k}$ the random weight $\tilde{M}_{t,k}/N_{\text{tot}}$, which is then multiplied by the additional weight $\tilde{w}_{t,k}$. Hence, $x_{t,k}$ receives the total weight

$$w'_{t,k} = \tilde{w}_{t,k} \frac{\tilde{M}_{t,k}}{N_{\text{tot}}}.$$  \hfill (B2)

If $m_{t,k} = j$, it holds that

$$E(w'_{t,k} | N_{t,j}) = \tilde{w}_{t,k} E(\frac{\tilde{M}_{t,k}/N_{\text{tot}}}{N_{t,j}} | N_{t,j})$$

$$= \frac{\pi_{t,j}}{N_{t,j}} \frac{w_{t,k} N_{t,j}}{\pi_{t,j}} = w_{t,k},$$  \hfill (B3)

i.e., on average the new weights $w'_{t,k}$ are equal to the original weights.

Appendix C: Temporal evolution of the mass balance state with the example of Findelgletscher

Figure C1. Temporal evolution of the ensemble mass balance state at stations FIN 1 and FIN 2. In the top two panels, the evolution of the mean and standard deviation of the filter (black lines and yellow shaded area) around the centered observations (blue lines and blue shaded area) is shown. In the bottom panel the mean deviation of the filter from the observations at both stations is shown.
Code and data availability. The camera observations are available under the following DOI: https://doi.org/10.3929/ethz-b-000508515 (Landmann, 2021). The meteorological data can be obtained as a paid service from https://www.meteoschweiz.admin.ch/home/klima/schweizer-klima-im-detail/raumliche-klimaanalysen.html (last access: 28 October 2021), and the glacier outlines and mass balances are available free of charge from the GLAMOS web site at https://doi.glamos.ch/data/inventory/inventory_sgi2010_r2010.zip (last access: 28 October 2021) and https://doi.glamos.ch/data/massbalance/massbalance_observation_elevationbins.csv (last access: 28 October 2021). The code used to produce results and figures can be obtained from the authors upon request.

Video supplement. Time lapse videos of all camera observations used in this study are available as videos under the following DOIs: PLM 1: https://doi.org/10.5446/48826 (Landmann, 2020a); FIN 1: https://doi.org/10.5446/48824 (Landmann, 2020b); FIN 2: https://doi.org/10.5446/48825 (Landmann, 2020c); RHO 1: https://doi.org/10.5446/48820 (Landmann, 2020d); RHO 2: https://doi.org/10.5446/48821 (Landmann, 2020e); RHO 3: https://doi.org/10.5446/48822 (Landmann, 2020f); RHO 4: https://doi.org/10.5446/48823 (Landmann, 2020g).

Author contributions. JML had the particle filter idea, implemented all models, did all figures, and wrote the paper. HRK supervised the particle filter methodology, brought in the method to prevent models from disappearing from the ensemble, and reviewed the paper. MH commented on the method, reviewed the paper, and mounted some of the stations. CO prepared and mounted most of the stations. MK commented on the particle filter and reviewed the paper. DF did the overall supervision, proposed to use data assimilation in JML’s doctorate, commented on the method, reviewed the paper, and acquired the funding.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We would like to acknowledge the funding that we got from GCOS (Global Climate Observing System) Switzerland and the extensive support that we got from the manufacturer of the cameras and transmitter boxes, Holfuy Ltd (in particular Gergely Mátyus). We would like to thank the teachers of the Joint ECMWF and University of Reading Data Assimilation Training Course that helped to significantly improve Johannes Marian Landmann’s knowledge on data assimilation, in particular Javier Amezua. Further, we would like to thank Anastasia Sycheva and Emmy Stigter for test-reading the methods chapter for reader friendliness. We appreciate the help from all people that conducted the field work apart from the authors, namely Malgorzata Chmiel, Amaury Dehecq, Lea Geibel, Katja Henz, Serafine Kattus, Johanna Klahold, Claudia Kurzböck, Amandine Sergeant, and Michaela Wenner, and we thank all people that took part in the round-robin experiment apart from the authors, namely Amaury Dehecq, Eef van Dongen, Elias Hodel, Jane Walden, and Michaela Wenner. Kudos to Dominik Gräff for helping to “TCify” Sect. 3.4 after the review comment by Douglas Brinkerhoff. We would also like to thank Bertrand Cluzet and colleagues for having changed the acronym of their project which coincidentally was the same as ours (CRAMON).

Review statement. This paper was edited by Marie Dumont and reviewed by Douglas Brinkerhoff and one anonymous referee.

References


Euronews: From Siberia to Switzerland, scorching August leads to more fires, less ice, available at: https://bit.ly/2BLLrU (last access: 28 October 2021), 2019.


